Data Integration Fundamentals

by Dave Wells

Module 0. About the Course (5 min)

Module 1. Data Integration Concepts (44 min)
- *Overview*
- *Data Integration Defined*
  - What is Data Integration
  - What isn't Data Integration
- *Data Integration Dependent Programs*
  - Common Data Integration Programs
  - Program vs. Project
  - Data Warehousing
    - Data Warehousing Defined
    - Data Warehouse Characteristics
    - ETL in Data Warehousing
    - Stages
    - Activities at Each Stage
  - Business Intelligence
    - BI Definition
    - BI Stack and Best Practices
  - Business Performance Management
    - BPM Defined
    - BPM Scope
    - BPM Cycle
  - Master Data Management
    - MDM Defined
    - How MDM Works
  - Operational Data Integration
    - Operational Data Store
    - Role of ODS
- *Data Integration Projects*
    - Program-Driven Projects
    - Application-Driven Projects

Module 2.—Data Integration Methods (49 min)

- *Overview*
- *Data Integration Techniques*
  - Data Consolidation
  - Data Virtualization
  - Data Federation
  - Data Synchronization
- *Data Integration Frequency*
  - One Time Integration
  - One Time Recurring Integration
  - Periodic Integration
  - Real-time Integration
- *Data Integration Systems*
  - Data Integration in the Technology Stack
  - Data Integration vs. Application Integration
  - Data Consolidation with ETL

  - o ETL Variations
  - o Extract decisions with ETL
  - o Data Virtualization
  - o Hybrid Solutions
  - o Mix and Match
  - o Other Integration System Functions
- *Data Integration Challenges*
  - o Understanding Data
  - o Getting Data
  - o Transforming Data
  - o Delivering Data
  - o Using Data
- *Data Integration Activities*
  - o Planning and Analysis
  - o Design and Construction
  - o Implementation and Execution

Module 3.—Understanding Data (61 min)
- *Overview*
- *Identifying Data Sources*
  - o Data Sources for Data Warehousing and BI
  - o Data Sources for ODS
  - o Data Sources for BPM
  - o Data Sources for MDM
  - o Data Sources for new System Startup
  - o Data Sources for Business Analytics
  - o Big Data as Sources Data
- *Profiling Data*
  - o Data Profiling Defined
  - o Column Profiling
  - o Table Profiling
  - o Cross-Table Profiling
  - o Unstructured Data Profiling
  - o Analyzing Data Profiles
  - o Using Data Profiles
- *Qualifying Data Sources*
  - o Usability
  - o Manageability
  - o Utility
- *Documenting Data Sources*
  - o Data Names
  - o Synonyms and Aliases
  - o Data Definitions
  - o Business Rules
  - o Data Models

Module 4. Integrating Data (78 min)
- *Overview*
- *Integration Requirements*

- o Unification and Rationalization
- o Case Study Introduction
- o How Much Unification
- o Entity Unification
- o Relationship Unification
- o Attribute Unification
- o Logical Data Model
- o Aggregation and Summary
- o Affinity Aggregation
- o Dimensional Aggregation
- o Point In Time Aggregation
- o Span of Time Aggregation
- o Content Aggregation
- o Aggregation: Logical Data Model Implications
- o Data Structures
- o Freshness and Latency
- o Quality and Usability
- *Data Capture*
  - o Data Capture Analysis and Design
  - o Source/Target Mapping
  - o Entity/Source/Target Mapping
  - o Data Element Mapping
  - o Data Capture Requirements
  - o Data Capture Design
  - o Data Capture Techniques
- *Data Transformation*
  - o Reasons to Transform Data
  - o Rule Examples
  - o Rule Dependencies
- *Data Delivery*
  - o Load Files
  - o Views
  - o Services

Module 5. Data Integration Architecture (37 min)

- *Overview*
- *Architecture Concepts*
  - o Architecture Purpose
  - o Components of Data Integration Architecture
  - o Integration Architecture Defined
  - o Data Sources, Middleware, and Data Consumers
- *Data Warehousing Architecture*
  - o Data Warehousing: What and Why?
  - o Common Data Warehousing Architectures
  - o Hub and Spoke with Top Down Development
  - o Bus with Bottom Down Development
  - o Hybrid Architecture with Blended Methodology
- *MDM Architecture*
  - o MDM: What and Why?
  - o Master Data Repository

- o Master Data Registry
- o MDM Engine
- o MDM Broker
- *Operational Data Integration Architecture*
  - o Operational Integration: What and Why?
  - o Data Integration vs. Application Integration
  - o Operational Data Store
  - o Data Synchronization

Module 6. Data Integration Roles and Responsibilities (28 min)

- *Overview*
- *Knowledge and Skills*
  - o Business, Data, and Technical Knowledge
  - o Skills Through the Data Integration Stages
- *Understanding the Data*
  - o Planning and Analysis
  - o Design and Construction
  - o Implementation and Execution
- *Getting the Data*
  - o Planning and Analysis
  - o Design and Construction
  - o Implementation and Execution
- *Transforming the Data*
  - o Planning and Analysis
  - o Design and Construction
  - o Implementation and Execution
- *Delivering the Data*
  - o Planning and Analysis
  - o Design and Construction
  - o Implementation and Execution
- *Using the Data*
  - o Planning and Analysis
  - o Design and Construction
  - o Implementation and Execution